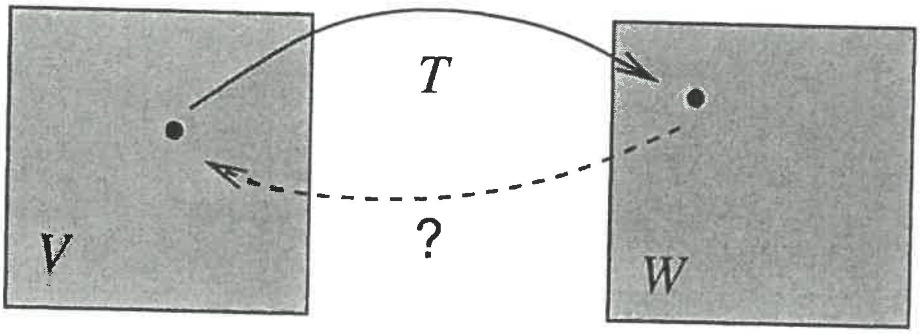
Invertibility

We would like to know when it is possible to "undo" a linear transformation.



To undo the action $\mathbf{v} \mapsto T(\mathbf{v})$ we need:

- Every $\mathbf{w} \in W$ must be the "image" of some $\mathbf{v} \in V$; *ie.* $\mathbf{w} = T(\mathbf{v})$,
- if $T(\mathbf{v}_1) = \mathbf{w} = T(\mathbf{v}_2)$ then we would like to know that $\mathbf{v}_1 = \mathbf{v}_2$ (otherwise, we would not know which $\mathbf{v}_i$ it was that $\mathbf{w}$ came from).

We formalize this intuitive idea in a definition:

Definition. A linear transformation $T : V \to W$ is called *onto* if $\mathcal{R}(T) = W$ and *one–to–one* (written 1–1) if

$$T(\mathbf{v}_1) = T(\mathbf{v}_2) \qquad \Rightarrow \qquad \mathbf{v}_1 = \mathbf{v}_2. \qquad \square$$

These two criteria essentially give "existence" and "uniqueness" of the "undo T" operation. Before giving some examples, we note that there are several equivalent conditions for 1–1:

Theorem 2.5 *Let $T : V \to W$ be a linear transformation. The following are equivalent:*
(i) T is 1–1;
(ii) $T(\mathbf{v}) = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$;
(iii) $\ker(T) = \{\mathbf{0}\}$;
(iv) if $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ are linearly independent in V then $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_r)\}$ are linearly independent in W;
(v) if B and B' are bases for V and W (respectively) then the equation $[T]_{B',B}\mathbf{x} = [\mathbf{w}]_{B'}$ has at most one solution.

Proof: (i)⇒(ii) By Theorem 2.1, $T(\mathbf{0}) = \mathbf{0}$. Thus, if T is 1–1 and $T(\mathbf{v}) = \mathbf{0}$ then $\mathbf{v} = \mathbf{0}$.
(ii)⇒(iii) Follows immediately from the definition of $\ker(T)$.
(iii)⇒(iv) Suppose that $\mathbf{0} = \sum_{i=1}^{r} a_i T(\mathbf{v}_i) = T(\sum_{i=1}^{r} a_i \mathbf{v}_i)$. Then $\sum_{i=1}^{r} a_i \mathbf{v}_i \in \ker(T) = \{\mathbf{0}\}$ so $\sum_{i=1}^{r} a_i \mathbf{v}_i = \mathbf{0}$. Since $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly independent, we have $a_i = 0$.
(iv)⇒(v) Let $B = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ be a basis for V. Then B is linearly independent, so $\{T(\mathbf{v}_1) \dots, T(\mathbf{v}_r)\}$ is linearly independent in B'. Thus, the matrix $[T]_{B',B}$ has linearly independent columns. Now suppose that $[T]_{B',B}\mathbf{x}_1 = [\mathbf{w}]_{B'} = [T]_{B',B}\mathbf{x}_2$. Then $[T]_{B',B}(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$, so $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0}$ by

the linear independence of the columns of $[T]_{B',B}$. Thus $\mathbf{x}_1 = \mathbf{x}_2$.
(v)⇒(i) Let $T(\mathbf{v}_1) = T(\mathbf{v}_2) = \mathbf{w}$. Then

$$[T]_{B',B}[\mathbf{v}_1]_B = [T(\mathbf{v}_1)]_{B'} = [\mathbf{w}]_{B'} = [T(\mathbf{v}_2)]_{B'} = [T]_{B',B}[\mathbf{v}_2]_B.$$

Thus, by (v), $[\mathbf{v}_1]_B = [\mathbf{v}_2]_B$ and hence $\mathbf{v}_1 = \mathbf{v}_2$. □

Examples

1. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$T(\mathbf{v}) = \begin{bmatrix} 1 & 1 & 2 \\ 1 & -1 & 1 \end{bmatrix} \mathbf{v}.$$

   Then T is onto but not 1–1.

2. Let $T_x : P_n \to P_{n+1}$ be the "multiply by x" operation:

$$T_x : p(x) \mapsto xp(x).$$

   Then T_x is 1–1 (as is easily checked), but not *onto* as there is no polynomial p such that $xp(x) = 1$.

3. Let $T_1 : P_n \to P_{n-1}$ be the operation of differentiation. Then T_1 is *onto*, but not 1–1.

4. For any vector space id is both 1–1 and onto.

We are now ready to say precisely what we mean by "undoing T".

Definition. If $T : V \to W$ is both onto and 1–1 then we will say that it is *invertible*. Then, there exists a transformation $T^{-1} : W \to V$ such that

$$\begin{aligned} T^{-1}(T(\mathbf{v})) &= \mathbf{v} \text{ for all } \mathbf{v} \in V \\ T\left(T^{-1}(\mathbf{w})\right) &= \mathbf{w} \text{ for all } \mathbf{w} \in W. \end{aligned}$$

T^{-1} is the transformation that allows us to "undo T". □

Invertible transformations are particularly nice since they essentially provide "perfect" structure preserving mappings between vector spaces. The onto condition guarantees that T "covers" all of W, and the 1–1 condition guarantees that no structure is lost. Indeed, we will regard two vector spaces which are related by an invertible linear transformation as essentially the same.

We will conclude this section with some statements about invertible transformations. We will later see how to find the matrix representation of an inverse transformation.

Theorem 2.6 *Suppose that $T : V \to W$ is an invertible linear transformation. Then*
(i) $T^{-1} : W \to V$ is an invertible linear transformation;
(ii) $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent in V if and only if $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_r)\}$ is linearly independent in W.

Proof: (i) That T^{-1} is an invertible transformation follows immediately from the definition: $(T^{-1})^{-1} = T$. One only needs to check linearity of T^{-1}:

$$\begin{aligned} T^{-1}(a\,\mathbf{v} + b\,\mathbf{w}) &= T^{-1}\left(a\,T\left[T^{-1}(\mathbf{v})\right] + b\,T\left[T^{-1}(\mathbf{w})\right]\right) \\ &= T^{-1}\left(T\left[a\,T^{-1}(\mathbf{v}) + b\,T^{-1}(\mathbf{w})\right]\right) \\ &= a\,T^{-1}(\mathbf{v}) + b\,T^{-1}(\mathbf{w}). \end{aligned}$$

(ii) Since T is invertible, it is 1–1, so the linear independence of $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ implies the linear independence of $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_r)\}$ by Theorem 2.5 ((i)≡(iv)). For the other direction, since T is invertible, T^{-1} is a well defined, 1–1 linear transformation so the linear independence of $\{T(\mathbf{v}_i)\}_{i=1}^{r}$ implies the linear independence of $\{T^{-1}[T(\mathbf{v}_i)]\}_{i=1}^{r} = \{\mathbf{v}_i\}_{i=1}^{r}$. □

Example. An $m \times n$ matrix taking $\mathbb{R}^n$ to $\mathbb{R}^m$ by

$$\mathbf{x} \mapsto A \times \mathbf{x}$$

is an invertible transformation only if $n = m$ and the matrix A is invertible. □

Isomorphism

Linear transformations are the homomorphisms (structure preserving mappings) between vector spaces. We will say that two vector spaces V and W are *isomorphic* if and only if there is an invertible linear transformation mapping V into W.

Theorem 2.7 *Two vector spaces V and W are isomorphic if and only if* $\dim(V) = \dim(W)$.

Proof: (⇒) Let $T : V \to W$ be an invertible linear transformation, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for V. Since every basis is linearly independent, Theorem 2.6 implies that $\{T(v_1), \ldots, T(v_r)\}$ is linearly independent in W. Thus, any basis for W contains at least r vectors, so $\dim(W) \geq \dim(V)$. A similar argument gives the reverse inequality, so the only possibility is equality of dimensions.
(⇐) Let $\dim(V) = \dim(W) = r$, let $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for V and $\{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$ a basis for W. Put

$$T(a_1\,\mathbf{v}_1 + \cdots + a_r\,\mathbf{v}_r) = a_1\,\mathbf{w}_1 + \cdots + a_r\,\mathbf{w}_r.$$

Then $T : V \to W$ is an onto linear transformation. Since $\{\mathbf{w}_i\}_{i=1}^{r}$ is a basis for W, $T(a_1\,\mathbf{v}_1 + \cdots + a_r\,\mathbf{v}_r) = \mathbf{0}$ only if $a_1 = a_2 = \cdots = a_r = 0$. This shows that $\ker(T) = \{\mathbf{0}\}$, so T is 1–1 (by Theorem 2.5). Thus, T is invertible, so V and W are isomorphic. □

Composition of linear transformations

We would like to be able to find the matrix representation of an inverse linear transformation. To help us do this, we will first make a brief study of the notion *composition*.

Definition. If $T_1 : V \to W$ and $T_2 : W \to U$ are linear transformations, we can define their *composition* $T_2 \circ T_1 : V \to U$ by

§ III ∘ Inner product spaces (over $\mathbb{C}$)

We will now add some additional geometric structure to vector spaces: an *inner product*. This will allow us to define such notions as orthogonality and "best approximations".

3.1 Inner product spaces

The standard basis for $\mathbb{R}^n$ is particularly nice to work with because the coordinate representation of any n–tuple of numbers can easily be read off. That is, if $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ denotes the standard basis, the vector $\mathbf{v} = (v_1, \ldots, v_n)$ has coordinate representation:

$$[\mathbf{v}]_B = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

For other bases, things may not be so simple:

Example. Let B be the standard basis for $\mathbb{R}^3$ and let $B' = \{(1,-1,2),(1,3,-2),(1,1,1)\}$ be another basis. Then, if $\mathbf{v} = (2,0,-1)$ we have

$$[\mathbf{v}]_B = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} \text{ whereas } [\mathbf{v}]_{B'} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}.$$

The representation with respect to B is easier! □

We would like to capture what it is about B that is so convenient; we will do this via the notion of *orthogonality*.

Definition. Let V be a vector space. An operation $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ ($\mathbb{R}$) is called an *inner product* on V if for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a \in \mathbb{C}$ ($\mathbb{R}$):

IP1 $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$;

IP2 $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$;

IP3 $\langle a\mathbf{u}, \mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle$;

IP4 $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

A vector space equipped with an inner product is called an *inner product space* (IPS).

Remark. We have defined complex and real inner product spaces together. The only simplification for real IPSs is that since every real number is its own conjugate, the axiom (IP1) becomes $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$. □

Examples

1. The most familiar example of an inner product space is the so–called "dot product" on $\mathbb{R}^n$:

$$\mathbf{u} \cdot \mathbf{v} = (u_1, \ldots, u_n) \cdot (v_1, \ldots, v_n) = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

On $\mathbb{C}^n$ the *standard inner product* is

$$\langle \mathbf{u}, \mathbf{v} \rangle = (u_1, \ldots, u_n) \cdot (v_1, \ldots, v_n) = u_1 \overline{v_1} + u_2 \overline{v_2} + \cdots + u_n \overline{v_n}.$$

We have to take the conjugates on $\mathbf{v}$ to satisfy (IP1).

2. There are many other inner products on $\mathbb{R}^n$, for example:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle (u_1, u_2), (v_1, v_2) \rangle \triangleq 3u_1 v_1 + u_1 v_2 + v_1 u_2 + 0.5 u_2 v_2$$

defines an inner product on $\mathbb{R}^2$.

3. Recall that $C(-1, 1)$ is the space of real valued continuous functions on the interval $(-1, 1)$. Then

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_{-1}^{1} \mathbf{f}(x)\, \mathbf{g}(x)\, dx$$

is an inner product.

Proof: (IP1) Follows immediately from the definition of $\langle \cdot, \cdot \rangle$ since $\mathbf{f}(x)\mathbf{g}(x) = \mathbf{g}(x)\mathbf{f}(x)$ for all $x \in (-1, 1)$. (IP2) and (IP3) follow from the linearity of integration:

$$\begin{aligned} \langle \mathbf{f} + \mathbf{g}, \mathbf{h} \rangle &= \int_{-1}^{1} (\mathbf{f}(x) + \mathbf{g}(x))\mathbf{h}(x)\, dx \\ &= \int_{-1}^{1} \mathbf{f}(x)\mathbf{h}(x)\, dx + \int_{-1}^{1} \mathbf{g}(x)\mathbf{h}(x)\, dx \\ &= \langle \mathbf{f}, \mathbf{h} \rangle + \langle \mathbf{g}, \mathbf{h} \rangle \end{aligned}$$

and

$$\langle a\mathbf{f}, \mathbf{g} \rangle = \int_{-1}^{1} a\mathbf{f}(x)\mathbf{g}(x)\, dx = a \int_{-1}^{1} \mathbf{f}(x)\mathbf{g}(x)\, dx = a\langle \mathbf{f}, \mathbf{g} \rangle.$$

For (IP4), note that

$$\langle \mathbf{f}, \mathbf{f} \rangle = \int_{-1}^{1} (\mathbf{f}(x))^2\, dx$$

is the integral of a positive function, and is thus non-negative. Moreover, the integral of $\mathbf{f}^2$ can only be 0 if $\mathbf{f}^2 = \mathbf{0}$.

4. Any finite-dimensional vector space is an inner product space.

Proof: Let V have dimension d. By Theorem 2.7, there is an invertible linear transformation $T : V \to \mathbb{R}^d$ (or $\mathbb{C}^d$ if V is a complex IPS). Define

$$\langle \mathbf{u}, \mathbf{v} \rangle = T(\mathbf{u}) \cdot \overline{T(\mathbf{v})}.$$

Axioms (IP1)–(IP3) follow immediately by properties of the dot product and the linearity of T. For (IP4), $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow T(\mathbf{x}) \cdot T(\mathbf{x}) = 0 \Leftrightarrow T(\mathbf{x}) = \mathbf{0}$ in $\mathbb{R}^d$ (or $\mathbb{C}^d$). But T is an invertible linear transformation, so it is certainly 1–1 and hence $\mathbf{x} = \mathbf{0}$ whenever $T(\mathbf{x}) = \mathbf{0}$. □

Just as with the dot product on $\mathbb{R}^n$, we can use any inner product to define the length of a vector.

Definition. The length of a vector $\mathbf{u}$ is

$$\|\mathbf{u}\|_{\langle\cdot,\cdot\rangle} = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle}.$$

(The subscript $\langle\cdot,\cdot\rangle$ emphasizes the dependence of the length on the inner product. When working in a fixed inner product space, this subscript will sometimes be omitted.) □

Examples

1. If $\mathbf{u} = (u_1, u_2, u_3)$ is a vector in $\mathbb{R}^3$ and $\langle\cdot,\cdot\rangle$ is the usual dot product then

$$\|\mathbf{u}\|_{\langle\cdot,\cdot\rangle} = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle} = \sqrt{u_1^2 + u_2^2 + u_3^2}.$$

2. With the usual inner product on $\mathbb{R}^2$,

$$\|(3,4)\|_{\langle\cdot,\cdot\rangle} = \sqrt{3^2 + 4^2} = \sqrt{25} = 5.$$

Note. We can use the IP to define the distance between vectors $\mathbf{u}$ and $\mathbf{v}$:

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_{\langle\cdot,\cdot\rangle}.$$

□

Theorem 3.1 [Cauchy–Schwarz inequality.] *For any two vectors* $\mathbf{u}, \mathbf{v}$ *in an inner product space,*

$$\mathcal{R}\ell(\langle \mathbf{u}, \mathbf{v}\rangle) \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

Proof: If either $\mathbf{u}$ or $\mathbf{v}$ is the zero vector then the result reduces to "$0 = 0$". Otherwise, let $f : \mathbb{R} \to \mathbb{R}$ be function:

$$f(t) = \|\mathbf{u} + t\mathbf{v}\|^2_{\langle\cdot,\cdot\rangle} = \langle \mathbf{u} + t\mathbf{v}, \mathbf{u} + t\mathbf{v}\rangle.$$

By (IP4), $f(t) \geq 0$ for any $t \in \mathbb{R}$, and by direct calculation:

$$\begin{aligned} f(t) &= \langle \mathbf{u} + t\mathbf{v}, \mathbf{u} + t\mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u} + t\mathbf{v}\rangle + \langle t\mathbf{v}, \mathbf{u} + t\mathbf{v}\rangle \\ &= \overline{\langle \mathbf{u} + t\mathbf{v}, \mathbf{u}\rangle} + t\,\langle \mathbf{v}, \mathbf{u} + t\mathbf{v}\rangle \\ &= \overline{\langle \mathbf{u}, \mathbf{u}\rangle} + \overline{\langle t\mathbf{v}, \mathbf{u}\rangle} + t\overline{\langle \mathbf{u} + t\mathbf{v}, \mathbf{v}\rangle} \\ &= \overline{\langle \mathbf{u}, \mathbf{u}\rangle} + \bar{t}\overline{\langle \mathbf{v}, \mathbf{u}\rangle} + t\overline{\langle \mathbf{u}, \mathbf{v}\rangle} + t\overline{\langle t\mathbf{v}, \mathbf{v}\rangle} \\ &= \|\mathbf{u}\|^2 + t(\langle \mathbf{u}, \mathbf{v}\rangle + \overline{\langle \mathbf{u}, \mathbf{v}\rangle}) + t^2\,\|\mathbf{v}\|^2 \\ &= c + bt + at^2 \end{aligned}$$

where $a = \|\mathbf{v}\|^2$, $b = 2\mathcal{R}\ell(\langle \mathbf{u}, \mathbf{v}\rangle)$ and $c = \|\mathbf{u}\|^2$. Since f is a non-negative quadratic function, it can have at most one zero. Therefore, by the discriminant test,

$$0 \geq b^2 - 4ac = 4\left((\mathcal{R}\ell(\langle \mathbf{u}, \mathbf{v}\rangle))^2 - \|\mathbf{u}\|^2\|\mathbf{v}\|^2\right).$$

The result follows. □

3.2 Orthogonality

Definition. We will say that two vectors $\mathbf{u}, \mathbf{v}$ are *orthogonal* if

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0.$$

We will write $\mathbf{u} \perp \mathbf{v}$. □

Example. If $\langle \cdot, \cdot \rangle$ is the usual dot product on $\mathbb{R}^3$ then

$$(1, -1, 2) \perp (1, -1, -1)$$

since

$$\langle (1, -1, 2), (1, -1, -1) \rangle = 1^2 + (-1)^2 + 2 \cdot (-1) = 0.$$

□

Theorem 3.2 [Generalized Pythagoras' theorem] *Let* $\mathbf{u} \perp \mathbf{v}$. *Then* $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Proof: Since $\mathbf{u} \perp \mathbf{v}$ we have $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Then

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
&= \|\mathbf{u}\|^2 + (\langle \mathbf{u}, \mathbf{v} \rangle + \overline{\langle \mathbf{u}, \mathbf{v} \rangle}) + \|\mathbf{v}\|^2 \\
&= \|\mathbf{u}\|^2 + (0 + \overline{0}) + \|\mathbf{v}\|^2 \\
&= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.
\end{aligned}$$

This is a generalization of Pythagoras' theorem because any triangle with sides $\mathbf{u}, \mathbf{v}$ and hypotenuse $\mathbf{u}+\mathbf{v}$ is a right–angled triangle if $\mathbf{u} \perp \mathbf{v}$. □

Best approximation

We will now see how to make the "best" approximation to an arbitrary vector $\mathbf{v}$ by a vector from a given subspace.

Definition. A vector $\mathbf{w} \in W$ will be called the *best approximation to* $\mathbf{v}$ if

$$\|\mathbf{w} - \mathbf{v}\| \leq \|\mathbf{w}' - \mathbf{v}\|$$

for every $\mathbf{w}' \in W$. □

Note: Since $\|\mathbf{w} - \mathbf{v}\|^2 = \min_{\mathbf{w}' \in W} \|\mathbf{w}' - \mathbf{v}\|^2$, the best approximation is sometimes called "least squares". □

Theorem 3.3 *Let* W *be a subspace of* V. *Then* $\mathbf{w} \in W$ *is the best approximation to* $\mathbf{v} \in V$ *if and only if* $(\mathbf{w} - \mathbf{v}) \perp \mathbf{x}$ *for every* $\mathbf{x} \in W$.

Proof: ($\Rightarrow$) Suppose $\mathbf{w}$ is the best approximation to $\mathbf{v}$ and $\mathbf{x} \in W$. Then

$$||\mathbf{v}-\mathbf{w}||^2 \leq ||\mathbf{v}-(\mathbf{w}+\mathbf{x})||^2 = ||\mathbf{v}-\mathbf{w}||^2 - 2\mathcal{R}\ell\langle \mathbf{v}-\mathbf{w},\mathbf{x}\rangle + ||\mathbf{x}||^2,$$

so that

$$0 \leq -2\mathcal{R}\ell\langle \mathbf{v}-\mathbf{w},\mathbf{x}\rangle + \langle \mathbf{x},\mathbf{x}\rangle.$$

Now let

$$f(t) = -2\mathcal{R}\ell\langle \mathbf{v}-\mathbf{w},t\mathbf{x}\rangle + \langle t\mathbf{x},t\mathbf{x}\rangle = -2t\mathcal{R}\ell\langle \mathbf{v}-\mathbf{w},\mathbf{x}\rangle + t^2\langle \mathbf{x},\mathbf{x}\rangle.$$

Since $f(t) \geq 0$, f has at most one zero, and the discriminant test implies that $\mathcal{R}\ell\langle \mathbf{v}-\mathbf{w},\mathbf{x}\rangle = 0$. A similar quadratic (using $i\,t\mathbf{x}$ in place of $t\mathbf{x}$) can be used to show that $\mathcal{I}m\langle \mathbf{v}-\mathbf{w},\mathbf{x}\rangle = 0$. Therefore, $\langle \mathbf{v}-\mathbf{w},\mathbf{x}\rangle = 0$.
($\Leftarrow$) Suppose that $\mathbf{x} \perp (\mathbf{v}-\mathbf{w})$ for every $\mathbf{x} \in W$. Then, if $\mathbf{w}' \in W$ let $\mathbf{x} = \mathbf{w}-\mathbf{w}' \in W$, so by Theorem 3.2,

$$||\mathbf{w}'-\mathbf{v}||^2 = ||(\mathbf{w}'-\mathbf{w})+(\mathbf{w}-\mathbf{v})||^2 = ||\mathbf{x}-(\mathbf{w}-\mathbf{v})||^2 = ||\mathbf{x}||^2 + ||(\mathbf{w}-\mathbf{v})||^2.$$

It follows that $||\mathbf{w}'-\mathbf{v}||^2 \geq ||\mathbf{w}-\mathbf{v}||^2$, and hence $\mathbf{w}$ is the best approximation to $\mathbf{v}$ in W. □

The existence of a best approximation depends on the topology of the subspace W. Since this can be a rather delicate matter, we note that best approximations always exist in finite-dimensional subspaces.

Orthogonal complement

One of the main objectives in studying inner product spaces is to be able to make nice decompositions of vectors (and subspaces) into "orthogonal components". We will often be interested in the collection of vectors which are orthogonal to a given subspace.

Definition. Let W be a subspace of an IPS V. Then the *orthogonal complement* $W^\perp$ to W is

$$W^\perp = \{\mathbf{v} \in V : \mathbf{w} \perp \mathbf{v} \text{ for every } \mathbf{w} \in W\}.$$

So $W^\perp$ is the set of vectors which are orthogonal to every vector from $\mathbf{v}$. □

Example. Let $V = \mathbb{R}^3$ and let W be the xy–plane. Then $W^\perp$ is the z–axis.

Proof: The xy–plane is simply:

$$W = \{\mathbf{w} \in \mathbb{R}^3 : \mathbf{w} = (x,y,0) \text{ where } x,y \in \mathbb{R}\}.$$

Thus,

$$W^\perp = \{\mathbf{v} \in \mathbb{R}^3 : \langle \mathbf{w},\mathbf{v}\rangle = 0 \text{ for all } \mathbf{w} \in W\}.$$

Since both $(1,0,0)$ and $(0,1,0)$ are members of W,

$$\begin{aligned} 0 &= \langle (1,0,0),(v_1,v_2,v_3)\rangle = 1v_1 + 0v_2 + 0v_3 = v_1 \\ 0 &= \langle (0,1,0),(v_1,v_2,v_3)\rangle = 0v_1 + 1v_2 + 0v_3 = v_2 \end{aligned}$$

for any $\mathbf{v} = (v_1,v_2,v_3)$ in $W^\perp$. Thus, $v_1 = v_2 = 0$. On the other hand, for any $x,y,z \in \mathbb{R}$,

$$\langle (x,y,0),(0,0,z)\rangle = x0 + y0 + 0z = 0.$$

Thus $(0, 0, z) \perp \mathbf{w}$ for any $\mathbf{w} \in W$. It follows that

$$W^{\perp} = \{\mathbf{v} = (0, 0, z) : z \in \mathbb{R}\}.$$

That is, the z–axis. □

Example. Let $\mathbf{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$ be any non–zero vector, and let W be the line through the $\mathbf{0}$ in the direction of $\mathbf{v}$. Thus,

$$W = \{a\mathbf{v} : a \in \mathbb{R}\}.$$

Then

$$W^{\perp} = \{(x, y, z) \in \mathbb{R}^3 : 0 = \langle \mathbf{v}, (x, y, z)\rangle = v_1 x + v_2 y + v_3 z\}.$$

This is simply the formula for the plane with normal $\mathbf{v}$! □

Example. Let $V = \mathbb{R}^n$, $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ and $W = \text{span}\,(S)$. Then

$$W^{\perp} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{v}_i = 0 \text{ for each } \mathbf{v}_i \in S\}.$$

If each $\mathbf{v}_i$ has coordinate representation

$$\mathbf{v}_i = [a_{i1}, a_{i2}, \ldots, a_{in}],$$

let A be the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rn} \end{bmatrix}.$$

Then for any $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in W^{\perp}$,

$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{r1}x_1 + \cdots + a_{rn}x_n \end{bmatrix} = \begin{bmatrix} \langle \mathbf{v}_1, \mathbf{x}\rangle \\ \langle \mathbf{v}_2, \mathbf{x}\rangle \\ \vdots \\ \langle \mathbf{v}_r, \mathbf{x}\rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}.$$

Thus, $W^{\perp}$ is the solution space to a matrix equation: it is the homogeneous solution space of the matrix whose rows span W! □

We now prove some results about orthogonal complements:

Theorem 3.4 *For any subspace W, $W^{\perp}$ is a subspace.*

Proof: By Theorem 1.2, we need only check the closure axioms. So, suppose $\mathbf{u}, \mathbf{v} \in W^{\perp}$ and a is a scalar. Then, for any $\mathbf{w} \in W$

$$\langle \mathbf{w}, \mathbf{u} + \mathbf{v}\rangle = \langle \mathbf{w}, \mathbf{u}\rangle + \langle \mathbf{w}, \mathbf{v}\rangle = 0 + 0 = 0,$$

and

$$\langle \mathbf{w}, a\mathbf{u}\rangle = \bar{a}\langle \mathbf{w}, \mathbf{u}\rangle = \bar{a}\,0 = 0.$$

Thus, $\mathbf{u} + \mathbf{v}, a\mathbf{u} \in W^{\perp}$. □

Theorem 3.5 *(i) $W^\perp \cap W = \{\mathbf{0}\}$; and (ii) If W is finite dimensional then $(W^\perp)^\perp = W$.*

Proof: (i) If $\mathbf{x} \in W^\perp$, then $\mathbf{x} \perp \mathbf{w}$ for every $\mathbf{w} \in W$. Thus, if also $\mathbf{x} \in W$, we must have $\mathbf{x} \perp \mathbf{x}$. Thus, $0 = \langle \mathbf{x}, \mathbf{x} \rangle$. By (IP4) this happens precisely when $\mathbf{x} = \mathbf{0}$. For part (ii), notice that

$$(W^\perp)^\perp = \{\mathbf{u} : \mathbf{u} \perp \mathbf{v} \text{ for all } \mathbf{v} \in W^\perp\}.$$

But if $\mathbf{u} \in W$, then certainly $\mathbf{u} \perp \mathbf{v}$ for all $\mathbf{v} \in W^\perp$. So we must have $W \subset (W^\perp)^\perp$. To complete the proof, we need to show that $(W^\perp)^\perp \subset W$. Let $\mathbf{v} \in (W^\perp)^\perp$ and let $\mathbf{w} \in W$ be the best approximation to $\mathbf{v}$. (The existence of such a $\mathbf{w}$ is guaranteed when W is finite dimensional.) By Theorem 3.3, $\mathbf{x} \perp (\mathbf{v} - \mathbf{w})$ for every $\mathbf{x} \in W$, so $\mathbf{v} - \mathbf{w} \in W^\perp$. But $\mathbf{v} \in (W^\perp)^\perp$ and $\mathbf{w} \in W \subset (W^\perp)^\perp$ so also $\mathbf{v} - \mathbf{w} \in (W^\perp)^\perp$. Thus, by part (i) of the Theorem, $\mathbf{v} - \mathbf{w} \in (W^\perp)^\perp \cap W^\perp = \{\mathbf{0}\}$. It follows that $\mathbf{v} = \mathbf{w}$. This shows that $\mathbf{v} \in W$, so $(W^\perp)^\perp \subset W$, and the proof is complete. □

Theorem 3.6 *If W is finite dimensional and $W^\perp = \{\mathbf{0}\}$ then $W = V$.*

Proof: Since $W = (W^\perp)^\perp$ this is a simple calculation:

$$W = (W^\perp)^\perp = \{\mathbf{0}\}^\perp = \{\mathbf{v} : \langle \mathbf{v}, \mathbf{0} \rangle = 0\} = V$$

since $\langle \mathbf{v}, \mathbf{0} \rangle = 0$ for every $\mathbf{v} \in V$. □

Our aim now is to show that every subspace W provides an "orthogonal decomposition" of V into a sum of W and $W^\perp$.

Definition. Let W_1 and W_2 be two subspaces of a vector space V. Then the *direct sum* of W_1 and W_2 is

$$W_1 \oplus W_2 = \{\mathbf{v} \in V : \mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2 \text{ where } \mathbf{w}_1 \in W_1, \mathbf{w}_2 \in W_2\}.$$

Thus, each vector in $W_1 \oplus W_2$ is made by adding together a vector from W_1 and a vector from W_2. □

Example. Let $V = \mathbb{R}^2$, and let W_1 be the x–axis and W_2 be the y axis. Then $W_1 = \{(x, 0) : x \in \mathbb{R}\}$, $W_2 = \{(0, y) : y \in \mathbb{R}\}$ and

$$W_1 \oplus W_2 = \{\mathbf{v} : \mathbf{v} = (x, 0) + (0, y)\} = \{\mathbf{v} : \mathbf{v} = (x, y)\} = \mathbb{R}^2.$$

Thus, we might say that $\mathbb{R}^2 = \mathbb{R} \oplus \mathbb{R}$. □

We now prove the main decomposition theorem:

Theorem 3.7 *For any IPS V and subspace W, $W \oplus W^\perp = V$.*

Proof: If we can prove that $(W \oplus W^\perp)^\perp = \{\mathbf{0}\}$ then the result will follow from Theorem 3.6. So, we suppose that $\mathbf{u} \in (W \oplus W^\perp)^\perp$. Then

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \text{ for every } \mathbf{v} \in W \oplus W^\perp.$$

Now, since $\mathbf{0} \in W^\perp$, we know that

$$\mathbf{w} = \mathbf{w} + \mathbf{0} \in W \oplus W^\perp \text{ for any } \mathbf{w} \in W.$$

Thus,

$$\langle \mathbf{u}, \mathbf{w} \rangle = 0 \text{ for any } \mathbf{w} \in W$$

so that $\mathbf{u} \in W^\perp$. An identical argument shows that $\mathbf{u} \in W$. Thus, $\mathbf{u} \in W \cap W^\perp = \{\mathbf{0}\}$. Since $\mathbf{u} \in (W \oplus W^\perp)^\perp$ is arbitrary, this shows that $(W \oplus W^\perp)^\perp = \{\mathbf{0}\}$, and the theorem follows. □

We can use Theorem 3.7 to make a unique "orthogonal decomposition" of any vector in V.

Theorem 3.8 *Let $\mathbf{v} \in V$ and let W be a finite dimensional subspace of V. Then there exist unique vectors $\mathbf{w} \in W$ and $\mathbf{u} \in W^\perp$ such that*

$$\mathbf{v} = \mathbf{w} + \mathbf{u}.$$

Proof: Since $V = W \oplus W^\perp$, the existence of $\mathbf{w}$ and $\mathbf{u}$ follows immediately from the definition of $\oplus$. To see that they are unique, suppose that

$$\mathbf{w} + \mathbf{u} = \mathbf{v} = \mathbf{w}' + \mathbf{u}'$$

are two such representations of $\mathbf{v}$. Then

$$\mathbf{w} - \mathbf{w}' = \mathbf{u}' - \mathbf{u}.$$

Since W is a subspace containing $\mathbf{w}$ and $\mathbf{w}'$, we must have $\mathbf{w} - \mathbf{w}' \in W$. Similarly, $\mathbf{u}' - \mathbf{u} \in W^\perp$. Thus

$$\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w} \in W \cap W^\perp.$$

But $W \cap W^\perp = \{\mathbf{0}\}$ by Theorem 3.5 (i). Thus $\mathbf{u} - \mathbf{u}' = \mathbf{0} = \mathbf{w} - \mathbf{w}'$. Thus $\mathbf{u} = \mathbf{u}'$, $\mathbf{w} = \mathbf{w}'$ and the decomposition is unique. □

Orthogonal projection

We can now define the notion of *orthogonal projection* into a subspace. For any vector $\mathbf{v} \in V$ and subspace W, we can use Theorem 3.8 to write

$$\mathbf{v} = \mathbf{w} + \mathbf{u}$$

where $\mathbf{w} \in W$ and $\mathbf{u} \in W^\perp$. We will call $\mathbf{w}$ the *projection of $\mathbf{v}$ onto W* and denote it:

$$\mathbf{w} = \text{proj}_W \mathbf{v}.$$

The projection is very nice, because

$$\mathbf{v} - \text{proj}_W \mathbf{v} = \mathbf{w} + \mathbf{u} - \mathbf{w} = \mathbf{u} \in W^\perp,$$

and is thus orthogonal to every vector in W (including $\text{proj}_W \mathbf{v}$). Consequently, by Theorem 3.3:

Theorem 3.9 *Let* $\mathbf{v} \in V$ *and let* W *be a finite dimensional subspace. Then*

$$\text{proj}_W \mathbf{v}$$

is the best approximation to $\mathbf{v}$.

We will explore some of the applications of orthogonal projection next, as well as deriving a useful formula!

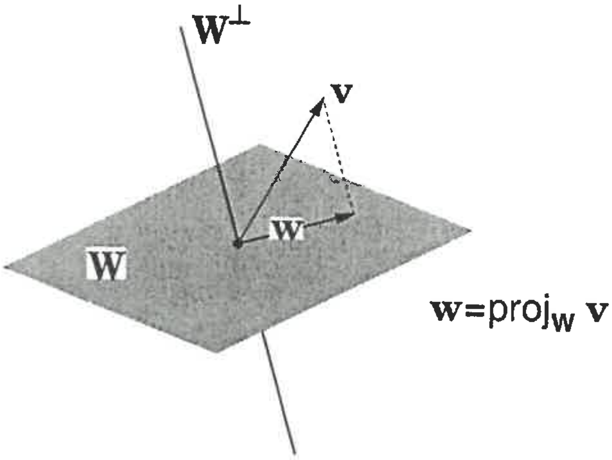


Figure. Diagram for orthogonal projection.

3.3 Formulas for projection

We will derive some formulas for orthogonal projection, and thus see why *orthogonal bases* are so desirable.

Recall that if $\mathbf{v} \in V$ then, for any subspace W of V,

$$\text{proj}_W \mathbf{v}$$

is a vector in W such that $\text{proj}_W \mathbf{v} \perp (\mathbf{v} - \text{proj}_W \mathbf{v})$; by Theorem 3.8 this is a unique decomposition of $\mathbf{v}$ into a pair of orthogonal vectors, one of which is in W.

Theorem 3.10 *Let* $W \subset V$ *be a subspace, and let* $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ *be a basis for* W. *Then*

$$\text{proj}_W \mathbf{v} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k$$

if and only if

$$\begin{pmatrix} \langle \mathbf{w}_1, \mathbf{w}_1\rangle & \langle \mathbf{w}_2, \mathbf{w}_1\rangle & \cdots & \langle \mathbf{w}_k, \mathbf{w}_1\rangle \\ \langle \mathbf{w}_1, \mathbf{w}_2\rangle & \langle \mathbf{w}_2, \mathbf{w}_2\rangle & \cdots & \langle \mathbf{w}_k, \mathbf{w}_2\rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{w}_1, \mathbf{w}_k\rangle & \langle \mathbf{w}_2, \mathbf{w}_k\rangle & \cdots & \langle \mathbf{w}_k, \mathbf{w}_k\rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = \begin{pmatrix} \langle \mathbf{v}, \mathbf{w}_1\rangle \\ \langle \mathbf{v}, \mathbf{w}_2\rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{w}_k\rangle \end{pmatrix}.$$

Proof: By Theorem 3.8 (uniqueness of decomposition), we have $c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k = \text{proj}_W\mathbf{v}$ if and only if

$$c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v} \in W^{\perp}.$$

We now show that this occurs if and only if the matrix equation is satisfied: the ith row of the matrix equation reads

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w}_i \rangle &= c_1\langle \mathbf{w}_1, \mathbf{w}_i \rangle + c_2\langle \mathbf{w}_2, \mathbf{w}_i \rangle + \cdots + c_k\langle \mathbf{w}_k, \mathbf{w}_i \rangle \\ &= \langle c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k, \mathbf{w}_i \rangle. \end{aligned}$$

Therefore, for each i,

$$\langle c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v}, \mathbf{w}_i \rangle = 0$$

so that

$$c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v} \perp \mathbf{w}_i$$

for each $\mathbf{w}_i$. ($\Rightarrow$) Now, if $c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v} \in W^{\perp}$, then certainly

$$c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v} \perp \mathbf{w}_i$$

for each $\mathbf{w}_i$. ($\Leftarrow$) On the other hand, if

$$\langle c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v}, \mathbf{w}_i \rangle = 0$$

then also

$$\left\langle c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v}, \sum_{i=1}^{k} a_i\mathbf{w}_i \right\rangle = 0$$

for any scalars a_i. Since every $\mathbf{w} \in W$ can be written as such a linear combination of $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$, we have

$$\langle c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v}, \mathbf{w} \rangle = 0$$

for every $\mathbf{w} \in W$. Thus

$$c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_k\mathbf{w}_k - \mathbf{v} \in W^{\perp}$$

and the theorem follows. □

Example. Let $V = \mathbb{R}^3$ with the standard basis, and let W be the xy–plane. So $W = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ where $\mathbf{e}_1, \mathbf{e}_2$ are standard basis vectors. Thus, if $v = (v_1, v_2, v_3)$,

$$\text{proj}_W\mathbf{v} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2$$

where

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \langle \mathbf{v}, \mathbf{e}_1 \rangle \\ \langle \mathbf{v}, \mathbf{e}_2 \rangle \end{pmatrix} = \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{e}_1 \rangle & \langle \mathbf{e}_2, \mathbf{e}_1 \rangle \\ \langle \mathbf{e}_1, \mathbf{e}_2 \rangle & \langle \mathbf{e}_2, \mathbf{e}_2 \rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Evaluating each of the inner products, we find that this equation reads

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Thus, $c_1 = v_1$ and $c_2 = v_2$. □

We can use this result to make projections onto lines:

Example. Let $V = \mathbb{R}^n$ and let W be a line in the direction $\mathbf{w}$:

$$W = \{a\mathbf{w} : a \in \mathbb{R}\}.$$

Then, for an arbitrary $\mathbf{v} \in \mathbb{R}^n$,

$$\text{proj}_W \mathbf{v} = c\mathbf{w}$$

if and only if

$$(\langle \mathbf{v}, \mathbf{w} \rangle) = (\langle \mathbf{w}, \mathbf{w} \rangle)(c).$$

Thus, $c = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{w}, \mathbf{w} \rangle}$. □

Example. Let W be any subspace of an inner product space V. Then the operation:

$$T(\mathbf{v}) = \text{proj}_W \mathbf{v}$$

is a linear transformation.

Proof: Let $\mathbf{u}, \mathbf{v} \in V$, and let $\mathbf{u}', \mathbf{v}' \in W^\perp$ be the unique vectors such that

$$\mathbf{u} = \text{proj}_W \mathbf{u} + \mathbf{u}' \text{ and } \mathbf{v} = \text{proj}_W \mathbf{v} + \mathbf{v}'$$

(recall Theorem 3.8). Then,

$$\text{proj}_W \mathbf{u} + \text{proj}_W \mathbf{v} \in W, \mathbf{u}' + \mathbf{v}' \in W^\perp$$

and

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= \text{proj}_W \mathbf{u} + \mathbf{u}' + \text{proj}_W \mathbf{v} + \mathbf{v}' \\ &= (\text{proj}_W \mathbf{u} + \text{proj}_W \mathbf{v}) + (\mathbf{u}' + \mathbf{v}'). \end{aligned}$$

Thus, by uniqueness of orthogonal projection,

$$\begin{aligned} T(\mathbf{u} + \mathbf{v}) &= \text{proj}_W (\mathbf{u} + \mathbf{v}) \\ &= (\text{proj}_W \mathbf{u} + \text{proj}_W \mathbf{v}) \\ &= T(\mathbf{u}) + T(\mathbf{v}). \end{aligned}$$

The proof that $T(c\mathbf{v}) = cT(\mathbf{v})$ is similar. □

Finally, Theorem 3.10 will let us do "change of basis":

Example. Let $V = \mathbb{R}^3$, and let $\mathbf{u}_1 = (1, 0, 1)$, $\mathbf{u}_2 = (1, 2, -1)$ and $\mathbf{u}_4 = (1, 2, 0)$, so that $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4\}$ is a basis. Regarding $W = \text{span}\,(S)$ as a subspace (even though it is all of $\mathbb{R}^3$), we can use projection to find the representation of a vector $\mathbf{v}$ with respect to S. By Theorem 3.10,

$$\mathbf{v} = \text{proj}_{\mathbb{R}^3} \mathbf{v} = \text{proj}_W \mathbf{v} = a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_4$$

if and only if

$$\begin{pmatrix} \langle \mathbf{v}, \mathbf{u}_1 \rangle \\ \langle \mathbf{v}, \mathbf{u}_2 \rangle \\ \langle \mathbf{v}, \mathbf{u}_4 \rangle \end{pmatrix} = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{u}_1 \rangle & \langle \mathbf{u}_2, \mathbf{u}_1 \rangle & \langle \mathbf{u}_4, \mathbf{u}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \langle \mathbf{u}_2, \mathbf{u}_2 \rangle & \langle \mathbf{u}_4, \mathbf{u}_2 \rangle \\ \langle \mathbf{u}_1, \mathbf{u}_4 \rangle & \langle \mathbf{u}_2, \mathbf{u}_4 \rangle & \langle \mathbf{u}_4, \mathbf{u}_4 \rangle \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

Thus, if $\mathbf{v} = (1, 0, 2)$, to find $[\mathbf{v}]_S$ we solve

$$\begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 6 & 5 \\ 1 & 5 & 5 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

The solution is $(a, b, c) = (1, -1, 1)$, so

$$\mathbf{v} = \mathbf{u}_1 - \mathbf{u}_2 + \mathbf{u}_4.$$

□

This method will obviously work for any basis. However, it is considerably easier in some bases than others. For example, in the standard basis, the matrix

$$A = \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{e}_1 \rangle & \langle \mathbf{e}_2, \mathbf{e}_1 \rangle & \cdots & \langle \mathbf{e}_n, \mathbf{e}_1 \rangle \\ \langle \mathbf{e}_1, \mathbf{e}_2 \rangle & \langle \mathbf{e}_2, \mathbf{e}_2 \rangle & \cdots & \langle \mathbf{e}_n, \mathbf{e}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{e}_1, \mathbf{e}_n \rangle & \langle \mathbf{e}_2, \mathbf{e}_n \rangle & \cdots & \langle \mathbf{e}_n, \mathbf{e}_n \rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

So the matrix equation is particularly easy to solve. Indeed, whenever the matrix A is *diagonal* it will be easy to read off a solution.

3.4 Applications of orthogonal projection

Example. Find the best approximation to $\mathbf{v} = (1, 0, 0)$ in span $(\mathbf{w}_1, \mathbf{w}_2)$ where

$$\mathbf{w}_1 = (1, 1, 1) \text{ and } \mathbf{w}_2 = (1, 0, 1).$$

Solution: By Theorem 3.9, the best approximation in W is $\text{proj}_W \mathbf{v}$. By Theorem 3.10, this can be found by solving the linear equation

$$\begin{aligned} \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} \langle \mathbf{w}_1, \mathbf{w}_1 \rangle & \langle \mathbf{w}_2, \mathbf{w}_1 \rangle \\ \langle \mathbf{w}_2, \mathbf{w}_1 \rangle & \langle \mathbf{w}_2, \mathbf{w}_2 \rangle \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \begin{pmatrix} \langle \mathbf{v}, \mathbf{w}_1 \rangle \\ \langle \mathbf{v}, \mathbf{w}_2 \rangle \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{aligned}$$

This equation has solution $a = 0, b = 1/2$. Thus,

$$\text{proj}_W \mathbf{v} = \frac{1}{2} \mathbf{w}_2 = (1/2, 0, 1/2).$$

The approximation error is

$$\|\mathbf{v} - \text{proj}_W \mathbf{v}\| = \|(1, 0, 0) - (1/2, 0, 1/2)\| = \sqrt{(1/2)^2 + 0^2 + (-1/2)^2} = \tfrac{1}{\sqrt{2}}.$$

□

Note: Best approximation is sometimes called "least squares" because

$$\|\text{proj}_W \mathbf{v} - \mathbf{v}\|^2 \leq \|\mathbf{w} - \mathbf{v}\|^2$$

for any $\mathbf{w} \in W$. Thus, $\text{proj}_W \mathbf{v}$ is the element of W for which the square of the *approximation error* $(\mathbf{w} - \mathbf{v})$ is least over all $\mathbf{w} \in W$. □

Application: fitting a trend line

Suppose we have a sequence of data points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

and we wish to approximate these by a straight line. Let us suppose that (x_i, y_i) are related by some unknown polynomial $f(x)$. We approximate f by a function of the form

$$y = A + Bx.$$

We can use best approximation. Let P_{n-1} be the space of degree-$(n-1)$ polynomials, and let an inner product be defined by:

$$\langle p(x), q(x) \rangle = p(x_1)q(x_1) + p(x_2)q(x_2) + \cdots + p(x_n)q(x_n).$$

We now compute the best approximation to f in span $(1, x)$. We are seeking a function of the form $A \times 1 + B \times x$:

$$\begin{pmatrix} \langle 1,1 \rangle & \langle x,1 \rangle \\ \langle 1,x \rangle & \langle x,x \rangle \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \langle f,1 \rangle \\ \langle f,x \rangle \end{pmatrix}.$$

Now, we do not know a formula for f, but we do know that

$$f(x_i) = y_i;$$

this is sufficient information to compute all of the inner products above:

$$\begin{aligned} \langle 1,1 \rangle &= 1 \times 1 + \cdots + 1 \times 1 = n, \\ \langle 1,x \rangle &= 1 \times x_1 + \cdots + 1 \times x_n = \sum_{i=1}^{n} x_i, \\ \langle x,x \rangle &= x_1 \times x_1 + \cdots + x_n \times x_n = \sum_{i=1}^{n} x_i^2, \\ \langle f,1 \rangle &= f(x_1) \times 1 + \cdots + f(x_n) \times 1 = \sum_{i=1}^{n} y_i, \\ \langle f,x \rangle &= f(x_1) \times x_1 + \cdots + f(x_n) \times x_n = \sum_{i=1}^{n} y_i x_i. \end{aligned}$$

Thus, the best fit to line to the data is given by the linear equation

$$y = A + Bx$$

where

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{pmatrix}.$$

3.5 Canonical decomposition with respect to a matrix

In this section we prove the "Fundamental Theorem of Linear Algebra"; that is, that associated to every linear transformation between vector spaces there is a natural orthogonal decomposition. To keep things simple, we will restrict attention to finite dimensional, real vector spaces. Consequently, every linear transformation can be represented by a real matrix. An $m \times n$ matrix A defines a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$ acting by matrix–vector multiplication: $\mathbf{x} \mapsto A\mathbf{x}$. Recall:

Definition. When A is a real $m \times n$ matrix,

$$\begin{aligned} \ker(A) &= \{\mathbf{x} \in \mathbb{R}^n \,|\, A\mathbf{x} = \mathbf{0}\}, \\ \mathcal{R}(A) &= \{\mathbf{w} \in \mathbb{R}^m \,|\, A\mathbf{v} = \mathbf{w} \text{ for some } \mathbf{v} \in \mathbb{R}^n\}. \end{aligned}$$

Similar definitions can be made for the transposed matrix A^T. □

Theorem 3.11 (Fundamental Theorem of Linear Algebra) *Let A be a real $m \times n$ matrix. Then, with the usual inner product, $(\mathcal{R}(A))^\perp = \ker(A^T)$. In particular, $\mathbb{R}^m = \ker(A^T) \oplus \mathcal{R}(A)$ and $\mathbb{R}^n = \ker(A) \oplus \mathcal{R}(A^T)$.*

Proof: In view of Theorem 3.7 we need to show that $(\mathcal{R}(A))^\perp = \ker(A^T)$. We do this by showing first that $\ker(A^T) \subset (\mathcal{R}(A))^\perp$ and then that $(\mathcal{R}(A))^\perp \subset \ker(A^T)$. Let $\mathbf{x} \in \ker(A^T)$ and $\mathbf{w} \in \mathcal{R}(A)$. Let $\mathbf{v} \in \mathbb{R}^n$ be such that $A\,\mathbf{v} = \mathbf{w}$. Then,

$$\langle \mathbf{x}, \mathbf{w} \rangle_{\mathbb{R}^m} = \mathbf{x} \cdot \mathbf{w} = \mathbf{x}^T\mathbf{w} = \mathbf{x}^T A\mathbf{v} = (A^T\mathbf{x})^T\mathbf{v} = \mathbf{0}^T\mathbf{v} = 0$$

since $\mathbf{x} \in \ker(A^T)$. This shows that $\mathbf{x} \perp \mathbf{w}$ for every $\mathbf{w} \in \mathcal{R}(A)$, so $\mathbf{x} \in (\mathcal{R}(A))^\perp$. Since $\mathbf{x} \in \ker(A^T)$ was arbitrary, we have established the first inclusion. For the other direction, let $\mathbf{x} \in (\mathcal{R}(A))^\perp$ and let $\mathbf{v} \in \mathbb{R}^n$ be arbitrary. Then $A\mathbf{v} \in \mathcal{R}(A)$, so $\mathbf{x} \perp A\mathbf{v}$ and

$$0 = \langle \mathbf{x}, A\mathbf{v} \rangle_{\mathbb{R}^m} = \mathbf{x}^T A\mathbf{v} = (A^T\mathbf{x})^T\mathbf{v} = \langle A^T\mathbf{x}, \mathbf{v} \rangle_{\mathbb{R}^n}.$$

This shows that $A^T\mathbf{x} \perp \mathbf{v}$ in $\mathbb{R}^n$ for *every* $\mathbf{v} \in \mathbb{R}^n$. In particular, $\langle A^T\mathbf{x}, A^T\mathbf{x} \rangle_{\mathbb{R}^n} = 0$, so $A^T\mathbf{x} = \mathbf{0}$ by (IP4). This establishes that $\mathbf{x} \in \ker(A^T)$, so that $(\mathcal{R}(A))^\perp \subset \ker(A^T)$ and the proof is complete. □

Some of the spaces in this theorem have special names:

$$\begin{aligned} \mathrm{col}(A) &= \mathrm{span}\,(\text{columns of } A) &= \mathcal{R}(A) \\ \mathrm{row}(A) &= \mathrm{span}\,(\text{rows of } A) &= \mathcal{R}(A^T) \\ \mathrm{null}(A) &= \ker(A). \end{aligned}$$

These are the column, row and null spaces of a matrix respectively. We can deduce directly from Theorem 3.11 that if A is a real $m \times n$ matrix then

$$\mathbb{R}^n = \mathrm{null}(A) \oplus \mathrm{row}(A) \text{ and } \mathbb{R}^m = \mathrm{null}(A^T) \oplus \mathrm{col}(A).$$

These decompositions allow us to prove the following surprising theorem, which justifies the hitherto mysterious fact that the *rank* of a matrix is well defined.

Theorem 3.12 *If A is an $m \times n$ matrix then* $\dim(row(A)) = \dim(col(A))$.

Proof: We really need to prove that $\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A^T))$. The result will follow from Theorem 2.7 if we can show that there is an invertible linear transformation mapping $\mathcal{R}(A^T)$ onto $\mathcal{R}(A)$. Let $\mathbf{v} \in \mathcal{R}(A^T)$ and put $T(\mathbf{v}) = A\mathbf{v}$. Clearly, $T : \mathcal{R}(A^T) \to \mathcal{R}(A)$; we need to show that T is 1–1 and onto. To see that T is onto, let $\mathbf{w} \in \mathcal{R}(A)$ and $\mathbf{y} \in \mathbb{R}^n$ be such that $A\mathbf{y} = \mathbf{w}$. By Theorems 3.7 and 3.11 there are vectors $\mathbf{x} \in \ker(A)$ and $\mathbf{v} \in \mathcal{R}(A^T)$ such that $\mathbf{y} = \mathbf{x} + \mathbf{v}$. Then $T(\mathbf{v}) = A\mathbf{v} = A(\mathbf{y} - \mathbf{x}) = A\mathbf{y} - A\mathbf{x} = \mathbf{w} - \mathbf{0} = \mathbf{w}$, so T is onto. To see that T is 1–1, suppose that $T(\mathbf{x}) = \mathbf{0}$ where $\mathbf{x} \in \mathcal{R}(A^T)$. Then, by definition, $\mathbf{x} \in \ker(A) = (\mathcal{R}(A^T))^\perp$. Hence, $\mathbf{x} = \mathbf{0}$ by Theorem 3.5 (i). It follows by Theorem 2.5 that T is 1–1. □

This theorem forms the theoretical basis of the method of using elementary row operations to solve matrix equations. Since the elementary row operations are invertible operations on the row space of a matrix, they do not change the dimension of the row space. Therefore, they do not change the dimension of the column space either. Since the main idea in solving linear systems is to find out how to write a given vector as a linear combination of the columns (think about what a matrix equation $A\mathbf{x} = \mathbf{b}$ is actually saying), it is very important to know that the dimension of the span of the columns is not altered by the solution process.

Remark. Theorem 3.11 can be extended to linear transformations between general IPSs. Then, if $T : V \to W$ is linear, we can define an *adjoint* transformation $T^* : W \to V$ by requiring

$$\langle T^*\mathbf{w}, \mathbf{v}\rangle_V = \langle \mathbf{w}, T(\mathbf{v})\rangle_W$$

for every $\mathbf{v} \in V$. Theorem 3.11 then states that W can be decomposed as $\ker(T^*) \oplus \mathcal{R}(T)$. □